



Ms. Marlene H. Dortch, Secretary  
Federal Communications Commission  
445 Twelfth Street, SW  
Washington, DC 20554

**July 29, 2009**

**Re: Notice of *Ex Parte* Presentation**

**RE: WC Docket No. 07-38, GN Docket No. 09-51**

Dear Ms. Dortch,

This letter is to advise you, in accordance with Section 1.1206(b) of the Commission's rules, that on July 28, 2009, Sascha Meinrath and Michael Calabrese of the New America Foundation with Blair Levin and John Leibovitz.

We discussed areas where current scientific research aligned with the Commission's research goals. Further, we clarified that we strongly support the Commission's efforts to improve data collection as well as the use of empirical research to inform policy-making. We presented concerns that the data necessary to conduct mission-critical Internet research was currently unavailable (see attached policy brief, "Analyzing in the Dark:

The Internet Research Data Acquisition Crisis

") and how the Commission could develop platforms for collecting useful data.

Sincerely,

Sascha Meinrath, Director  
Open Technology Initiative  
New America Foundation  
meinrath@newamerica.net

## Analyzing in the Dark: The Internet Research Data Acquisition Crisis

When the NSFnet backbone was privatized in 1995 the network science community lost access to the only set of publicly available statistics on a national Internet backbone network. This transition essentially eliminated the opportunity to conduct analyses on a widely-used backbone. Today, far from having an analytic handle on the Internet, network researchers often lack the ability to measure traffic at the granularity necessary to make increasingly critical infrastructure improvements. Legislators operate under an enforced ignorance of potential security problems when the scientific community is unable to identify potential congestion points on the internet, and empirically clueless about how the Internet can be improved. Access to network traffic data would allow researchers to begin solving problems of cyber-security, spam overload, privacy invasion and identity theft, digital rights management and piracy, network congestion, pricing discrimination, illegal pornography and a host of other issues. Without these data, network scientists, regulators, and decision-makers are left fumbling in the dark as we attempt to address these seemingly intractable and growing problems.

Over the past decade, while the core of the Internet has continued to expand, scientific measurement and modeling of its systemic characteristics has largely stalled. An inevitable problem with contemporary Internet traffic measurement studies is that they are quickly made obsolete in an environment where traffic, technology, and topology change faster than we can currently measure them. The proliferation of multimedia content and new services and applications makes the acquisition of data far more difficult and costly than in previous years. Much like a scanning electron microscope is a critical tool for modern physics laboratories, high-powered and expensive measurement tools are needed by Internet researchers to keep pace with the Internet's increasing complexity.

These problems have recently been identified by the Department of Homeland Security (DHS) as critically important to the longterm interests and security of the United States. Recently, DHS recognized the need to support the calibration of cyber-security tools in real world environments and has launched the PREDICT Project to allow researchers to request datasets to assist their research into cyber-defense technologies, products, models and strategies. DHS has facilitated progress in the legal and privacy facets of infrastructure data access, specifically addressing the concerns of Internet Service Providers who want to support the research community but are constrained by privacy laws or policies. However, while DHS is actively supporting the sharing of existing datasets for research and analysis, it has no budget for conducting research on the infrastructure itself. Meanwhile, the National Science Foundation has cut measurement infrastructure project budgets by 75% or more due to funding constraints, thus exacerbating the problem.

We cannot hope to build a national broadband policy that brings America into the digital future without a solid understanding of what is happening on our networks. Throughout the decades that the United States government was steward of the early Internet, the only statistics collected regularly were those required by government contract. Since the privatization of the Internet in the mid-1990s, we have embraced a policy that has sacrificed this data access, assuming that the less regulation of the Internet, the better. What is absolutely clear, however, is that this privatization has created disastrous outcomes for network science and basic research due to the lack of regulatory requirements for transparency. Because of the pervasiveness of non-disclosure agreements and the practice of treating even mundane operational practices as trade secrets, today's network science operates in a self-perpetuating "fog of unknowing" around the Internet.

The Internet Data Acquisition Crisis is multifaceted and includes numerous mission-critical elements that network researchers cannot measure. Taken together, these "unknowns" are creating clear national security concerns by undermining our knowledge about this critical information infrastructure. Notable topic areas researchers are particularly concerned about include:

- network topology from one point to another (in either direction, at any network layer);
- propagation of a routing update across the Internet (i.e., how robust routing is);
- core router information such as their Router Information Base (RIB) – researchers can only gain access to a router's Forwarding Information Base (FIB);
- precise one-way delay from two places on the Internet (i.e., how efficient the routing is);
- hour packet information in the Internet's core (e.g., the collection of packet traces, even anonymized, from any backbone available to any academic researcher);
- accurate flow counts from the Internet's core (i.e., how much throughput);
- anything from the Internet's core with real IP addresses (e.g., where data is coming from/to);
- the topology of the Internet's core (i.e., how does the Internet's core interconnect);
- accurate bandwidth or capacity information – not even along a path, much less for each link (i.e., link performance information);
- how much, in absolute or relative terms, spam/phishing/viruses/botnets/hosts/routers exists;
- information on what is causing problems (e.g., the existence of obsolete or misconfigured software, network and protocol parameters, bad cables, lossy media, insufficient send/receive buffers, route instability, VPNs, firewalls);
- cost structure data; and,
- potential privacy/legal issues in the Internet's operations.

Important questions researchers cannot answer due to this Data Acquisition Crisis:

- What percent of Internet users in the US (or world) are running p2p file sharing applications?
- What proportion of traffic going across a given network backbone is: spam, malware (botnets, worms, phishing), encrypted, or real-time? How fast are these categories growing? How much overhead do they represent on the network? How do they shift the economics of the network?
- What are the effects of outages, new routing policies, and other topology changes on surrounding (or distant) Internet Service Providers (ISPs)?
- Which providers control how much Internet topology/bandwidth resources in the US?
- Which segments of the infrastructure are especially vulnerable (or subject) to congestion?
- What is the extent of asymmetric routing and route instability as a function of ISP and over time? What is the impact of asymmetry on performance?
- What are the root causes of Internet outages, and how does the distribution change over time?
- How effectively are we utilizing IPv4 address space (projected to be used up in 3-5 years) and BGP routing table space?
- How efficiently does the Internet backbone move traffic around? What percent of traffic is unwanted by the destination receiving it? What artificial bottlenecks exist?

Critical research areas scientists cannot pursue because of this Data Acquisition Crisis:

- Sustainable interdomain routing and addressing architecture (particularly important since we will run out of IPv4 addresses within the next half-decade).
- Improvements in congestion control. It is widely recognized that the most prevalent data transport protocol in the Internet, TCP, is not behaving efficiently at modern bandwidths and for real-time applications such as voice.
- Measurement technology itself – it is difficult to justify investing resources into measurement technology when there is no promise of a network (or traffic, on a testbed) to measure.
- Innovative security, multimedia, and IP transport technologies (e.g., DNSSEC, S-BGP, multicast, RSVP/QOS).